GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:47:21 ; Search time 210.42 Seconds
(without alignments)
6.500 Million cell updates/sec

Title: US-09-331-631A-8_COPY_80_119
Perfect score: 225
Sequence: 1 PEDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	225	100.0	590	19	W62832	Gossypium hirsutum
2	119	52.9	525	19	W62831	Theobroma cacao an
3	119	52.9	566	13	R20181	Sequence encoded b
4	114	50.7	666	19	W62828	Macadamia integrif
5	112	49.8	625	19	W62830	Macadamia integrif
6	112	49.8	666	19	W62829	Macadamia integrif
7	82	36.4	371	20	W73369	Epitope tagged TBP
8	77.5	34.4	2023	21	Y54320	Amino acid sequenc
9	76.5	34.0	2074	21	Y54319	Amino acid sequenc
10	75	33.3	910	20	Y22191	Mouse brain CNG-1
11	73	32.4	86	20	W95073	GST-HD fusion prot
12	73	32.4	86	20	W95078	GST-HD fusion prot
13	73	32.4	94	20	W95075	GST-HD fusion prot
14	73	32.4	94	20	W95080	GST-HD fusion prot
15	72.5	32.2	1299	21	Y58633	Protein regulating
16	72	32.0	28	19	W62841	Stenocarpus sinuat
17	70	31.1	108	20	W95071	Amino acid sequenc
18	70	31.1	108	20	W95076	Amino acid sequenc
19	69	30.7	1162	21	Y58500	HHV8 ORF 73 protei
20	69	30.7	1326	20	Y55933	Human ZC3 protein.
21	67.5	30.0	609	19	W83215	Human h-NUMB-R. H
22	67	29.8	732	19	W63715	Rat rSK3 protein.
23	67	29.8	1898	20	Y30795	A human trichohyal
24	66	29.3	637	19	W62837	Hordeum vulgare an
25	66	29.3	1420	20	W81025	AIB1 (Amplified in
26	66	29.3	1522	20	Y21975	Human steroid rece
27	65.5	29.1	303	15	R60054	Dirofilaria immiti
28	65.5	29.1	404	18	W14909	Baker's yeast nega
29	65	28.9	360	20	Y33492	Human Machado-Jose
30	65	28.9	513	20	W88413	Acute myeloid leuk
31	65	28.9	548	20	W89189	Alternatively spli
32	65	28.9	596	20	W89184	Osf2/Cbfa1 native
33	65	28.9	816	16	R71111	Spinocerebellar at
34	65	28.9	816	20	Y33494	Human SCA1 protein
35	65	28.9	1138	21	Y83222	CAP6 polypeptide (
36	64.5	28.7	404	13	R27284	Sequence encoded b
37	64	28.4	919	10	P93109	Human androgen rec
38	64	28.4	919	18	W14783	Androgen receptor.
39	64	28.4	919	21	Y78914	Human androgen rec
40	63.5	28.2	154	20	Y33504	Human unliganded a
41	63.5	28.2	918	12	R12223	Human androgen rec
42	63.5	28.2	918	20	Y33491	Human androgen rec
43	63.5	28.2	1360	21	Y85263	Human protein kina
44	63	28.0	1447	20	W81029	Murine pCIP protei
45	62.5	27.8	314	21	Y91560	Human secreted pro

ALIGNMENTS

RESULT 1
W62832
ID W62832 standard; Protein; 590 AA.
XX
AC W62832;
XX
DT 27-OCT-1998 (first entry)
XX
DE Gossypium hirsutum antimicrobial protein.
XX
KW antimicrobial protein; infestation; control.
XX
OS Gossypium hirsutum.
XX
PN WO9827805-A1.
XX
PD 02-JUL-1998.
XX
PF 22-DEC-1997; 97WO-AU00874.
XX
PR 20-DEC-1996; 96AU-0004275.
XX
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
XX
DR WPI; 1998-377279/32.
XX
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals
XX
PS Claim 1; Page 49-51; 96pp; English.
XX
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian

```
CC   animals.
XX
SQ   Sequence   590 AA;


  Query Match            100.0%;  Score 225;  DB 19;  Length 590;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-17;
  Matches   40;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PEDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
          ||||||||||||||||||||||||||||||||||||||||
Db     80 pedpqrryeecqqecrqqeerqqpqcqqrclkrfeqeqqq 119


RESULT   2
W62831
ID   W62831 standard; Protein; 525 AA.
XX
AC   W62831;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Theobroma cacao antimicrobial protein.
XX
KW   antimicrobial protein; infestation; control.
XX
OS   Theobroma cacao.
XX
PN   WO9827805-A1.
XX
PD   02-JUL-1998.
XX
PF   22-DEC-1997;   97WO-AU00874.
XX
PR   20-DEC-1996;   96AU-0004275.
XX
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI   Bower NI,  Goulter KC,  Green JL,  Manners JM,  Marcus JP;
XX
DR   WPI; 1998-377279/32.
XX
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
XX
PS   Claim 1; Page 47-49; 96pp; English.
XX
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
XX
SQ   Sequence   525 AA;


  Query Match            52.9%;  Score 119;  DB 19;  Length 525;
  Best Local Similarity  45.0%;  Pred. No. 6.7e-06;
  Matches   18;  Conservative   14;  Mismatches    6;  Indels    2;  Gaps    1;

Qy      3 DPQRRYEECQQECRQQ--EERQQPQCQQRCLKRFEQEQQQ 40
          ||:::||:||: |  :  |||:| ||:||| : ::::|:|
Db     37 dprqqyeqcqrrceseateereqeqceqrcereykeqqrq 76


RESULT   3
R20181
ID   R20181 standard; Protein; 566 AA.
XX
AC   R20181;
XX
DT   16-APR-1992  (first entry)
XX
DE   Sequence encoded by 67 kD T. cacao protein cDNA.
XX
KW   Cocoa; flavour; vicilin; seed storage protein.
XX
OS   Theobroma cacao.
XX
PN   WO9119801-A.
XX
PD   26-DEC-1991.
XX
PF   07-JUN-1991;   91WO-GB00914.
XX
PR   11-JUN-1990;   90GB-0013016.
XX
PA   (MRSC ) MARS UK LTD.
XX
PI   Spencer ME,  Hodge R,  Deakin EA,  Ashton S;
XX
DR   WPI; 1992-024418/03.
DR   N-PSDB; Q20377.
XX
PT   Recombinant cocoa proteins - are responsible for flavour in cocoa
PT   beans and produced in large quantities using yeast and bacterial
PT   expression vectors
XX
PS   Claim 4; Fig 2; 59pp; English.
XX
CC   The inventors claim a 67 kD and 31 kD T. cacao protein, and
CC   fragments, and encoding DNAs.  The 47 kD and 31 kD proteins are
CC   derived from the 67 kD precursor.  T. cacao protein cDNA was
CC   detected in a cDNA library prepared from immature cocoa beans RNA
CC   using a probe based on the AA sequence of a CNBr peptide common to
CC   the 47 kD and 31 kD polypeptides.  Homology searches revealed close
CC   homologies between the 67 kD polypeptide and the vicilins, which are
CC   seed storage proteins.
XX
SQ   Sequence   566 AA;


  Query Match            52.9%;  Score 119;  DB 13;  Length 566;
  Best Local Similarity  45.0%;  Pred. No. 7.2e-06;
  Matches   18;  Conservative   14;  Mismatches    6;  Indels    2;  Gaps    1;

Qy      3 DPQRRYEECQQECRQQ--EERQQPQCQQRCLKRFEQEQQQ 40
          ||:::||:||: |  :  |||:| ||:||| : ::::|:|
Db     37 dprqqyeqcqrrceseateereqeqceqrcereykeqqrq 76


RESULT   4
W62828
ID   W62828 standard; Protein; 666 AA.
XX
AC   W62828;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Macadamia integrifolia antimicrobial protein.
XX
KW   antimicrobial protein; infestation; control.
XX
OS   Macadamia integrifolia.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..28
FT                   /note= "signal peptide"
FT   Protein         29..666
FT                   /note= "mature protein"
XX
PN   WO9827805-A1.
XX
PD   02-JUL-1998.
XX
PF   22-DEC-1997;   97WO-AU00874.
XX
PR   20-DEC-1996;   96AU-0004275.
```

```
XX
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI   Bower NI,  Goulter KC,  Green JL,  Manners JM,  Marcus JP;
XX
DR   WPI; 1998-377279/32.
DR   N-PSDB; V42310.
XX
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
XX
PS   Claim 1; Page 34-36; 96pp; English.
XX
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
XX
SQ   Sequence   666 AA;


  Query Match            50.7%;  Score 114;  DB 19;  Length 666;
  Best Local Similarity  47.5%;  Pred. No. 3e-05;
  Matches   19;  Conservative   11;  Mismatches    8;  Indels    2;  Gaps    1;

Qy       3 DPQRRYEECQQEC--RQQEERQQPQCQQRCLKRFEQEQQQ 40
           |||::||:||: |  |: | |    ||||| :|:|:|:::
Db     121 dpqqqyeqcqkhcqrreteprhmqtcqqrcerryekekrk 160


RESULT   5
W62830
ID   W62830 standard; Protein; 625 AA.
XX
AC   W62830;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Macadamia integrifolia antimicrobial protein.
XX
KW   antimicrobial protein; infestation; control.
XX
OS   Macadamia integrifolia.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..28
FT                   /note= "signal peptide"
FT   Protein         29..666
FT                   /note= "mature protein"
XX
PN   WO9827805-A1.
XX
PD   02-JUL-1998.
XX
PF   22-DEC-1997;   97WO-AU00874.
XX
PR   20-DEC-1996;   96AU-0004275.
XX
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI   Bower NI,  Goulter KC,  Green JL,  Manners JM,  Marcus JP;
XX
DR   WPI; 1998-377279/32.
DR   N-PSDB; V42316.
XX
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
XX
PS   Claim 1; Page 43-45; 96pp; English.
XX
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
XX
SQ   Sequence   625 AA;


  Query Match            49.8%;  Score 112;  DB 19;  Length 625;
  Best Local Similarity  47.5%;  Pred. No. 4.6e-05;
  Matches   19;  Conservative   11;  Mismatches    8;  Indels    2;  Gaps    1;

Qy       3 DPQRRYEECQQEC--RQQEERQQPQCQQRCLKRFEQEQQQ 40
           |||::||:||: |  |: | |    ||||| :|:|:|:::
Db     80 dpqqqyeqcqkrcqrreteprhmqicqqrcerryekekrk 119


RESULT   6
W62829
ID   W62829 standard; Protein; 666 AA.
XX
AC   W62829;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Macadamia integrifolia antimicrobial protein.
XX
KW   antimicrobial protein; infestation; control.
XX
OS   Macadamia integrifolia.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..28
FT                   /note= "signal peptide"
FT   Protein         29..666
FT                   /note= "mature protein"
XX
PN   WO9827805-A1.
XX
PD   02-JUL-1998.
XX
PF   22-DEC-1997;   97WO-AU00874.
XX
PR   20-DEC-1996;   96AU-0004275.
XX
PA   (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI   Bower NI,  Goulter KC,  Green JL,  Manners JM,  Marcus JP;
XX
DR   WPI; 1998-377279/32.
DR   N-PSDB; V42311.
XX
PT   Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT   useful for controlling microbial infestations of plants or mammals
XX
PS   Claim 1; Page 39-41; 96pp; English.
XX
CC   The sequence is that of an antimicrobial protein which can
CC   be used to control microbial infestations in plants and mammalian
CC   animals.
XX
SQ   Sequence   666 AA;


  Query Match            49.8%;  Score 112;  DB 19;  Length 666;
  Best Local Similarity  56.4%;  Pred. No. 4.9e-05;
  Matches   22;  Conservative    7;  Mismatches    4;  Indels    6;  Gaps    2;

Qy       3 DPQRR-YEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           |||:| ||:|:: | ||| ||| |||:||     :|||:
Db     188 dpqqreyedcrrrceqqeprqqyqcqrrc-----reqqr 221


RESULT   7
W73369
ID   W73369 standard; peptide; 371 AA.
XX
AC   W73369;
```

```
XX
DT   12-FEB-1999  (first entry)
XX
DE   Epitope tagged TBP protein.
XX
KW   TATA-box binding protein; epitope-tagged TBP; transcription complex; TAF;
KW   TBP associated factor; TAF-interaction factor; gene expression regulator.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   EP881288-A1.
XX
PD   02-DEC-1998.
XX
PF   26-MAY-1998;   98EP-0109516.
XX
PR   26-MAY-1997;   97EP-0108433.
XX
PA   (FARH ) HOECHST AG.
XX
PI   Berglund E,  Kirschbaum B,  Meisterernst M,  Polites G;
XX
DR   WPI; 1999-001394/01.
XX
PT   Transgenic animal expressing epitope-tagged TATA-box binding protein
PT   - for isolating higher-order transcription complexes and specific
PT   factors that associate with the protein, useful as potential
PT   therapeutic agents
XX
PS   Claim 5; Page 20-22; 38pp; English.
XX
CC   This sequence represents an epitope-tagged TATA-box binding protein (TBP)
CC   that is expressed by the transgenic non-human animals of the invention.
CC   The animals are used to produce TBP. TBP is used to isolate and
CC   characterise higher-order transcription complexes (from different tissue
CC   and cell types, optionally at different developmental stages). It is also
CC   used to identify new and/or specific TBP associated factors (TAFs,
CC   e.g. transcription factors, activators or inhibitors) and TAF-interaction
CC   factors, and to raise antibodies against TBP. The TAFs may be useful for
CC   regulating gene expression, e.g. disease-related genes, so are potential
CC   pharmaceuticals, also for identifying human analogues for use in drug
CC   screening. The antibodies are used for affinity purification of TBP and
CC   its complexes. TBP can isolate transcription complexes from a wide
CC   variety of different tissues and cells (contrast known methods that are
CC   limited to isolation from a particular cell type).
XX
SQ   Sequence   371 AA;

  Query Match             36.4%;  Score 82;  DB 20;  Length 371;
  Best Local Similarity   41.0%;  Pred. No. 0.056;
  Matches   16;  Conservative   14;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 EDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           |: ||: :: ||: :||:::|| | ||:  :: :|:|||
Db      85 eeqqrqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq 123


RESULT   8
Y54320
ID   Y54320 standard; Protein; 2023 AA.
XX
AC   Y54320;
XX
DT   06-APR-2000  (first entry)
XX
DE   Amino acid sequence of a human PCTG4 protein.
XX
KW   Human; PCTG4 region; X chromosome; q13 region; polymorphism;
KW   mental retardation; autism; depression; bipolar affective disorder;
KW   hypothyroidism; OPA gene; neuropsychiatric disorder.
XX
OS   Homo sapiens.
XX
PN   WO9955915-A2.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09365.
XX
PR   29-APR-1998;   98US-0083465.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA   (IOWA ) UNIV IOWA RES FOUND.
XX
PI   Philibert RA,  Ginns EI;
XX
DR   WPI; 2000-126357/11.
XX
PT   Identification of polymorphisms in the PCTG4 region of Xq13 for
PT   diagnosing mental retardation or autism -
XX
PS   Example 7; Page 81-84; 100pp; English.
XX
CC   The present sequence represents a human PCTG4 protein. Polymorphisms
CC   in the human PCTG4 region of chromosome Xq13 are associated with
CC   mental retardation, autism, depression, bipolar affective disorder or
CC   hypothyroidism. One 12 bp insertion polymorphism occurs within the
CC   coding region of the human OPA gene, and introduces a 4 amino acid
CC   insertion in a putative OPA domain. This domain has been shown to be
CC   involved in tissue specific expression. Another polymorphism consists
CC   of a pentanucleotide repeat approximately 7 kb upstream of the 12 bp
CC   polymorphism. Another polymorphisms consists of a dinucleotide repeat
CC   approximately 4.5 kb downstream of the 12 bp polymorphism. The
CC   specification describes a method for screening for polymorphisms in a
CC   PCTG4 nucleic acid sequence obtained from a subject. The PCTG4 related
CC   sequences within the q13 region of the X chromosome have polymorphisms
CC   associated with neuropsychiatric disorders.  The methods can be used to
CC   screen for the presence of a heritably linked form of mental retardation,
CC   autism, depression, bipolar affective disorder or hypothyroidism.
XX
SQ   Sequence   2023 AA;


  Query Match             34.4%;  Score 77.5;  DB 21;  Length 2023;
  Best Local Similarity   34.7%;  Pred. No. 0.87;
  Matches   17;  Conservative   14;  Mismatches    9;  Indels    9;  Gaps    1;

Qy       1 PEDPQRRYEECQQECRQQEERQQ---------PQCQQRCLKRFEQEQQQ 40
           ||  |:: :: ||: :||:::||          | ||: |:: :|:|||
Db    1896 peqqqqqqqqqqqqqqqqqqqqqqqqqqyhirqqqqqqilrqqqqqqq 1944


RESULT   9
Y54319
ID   Y54319 standard; Protein; 2074 AA.
XX
AC   Y54319;
XX
DT   06-APR-2000  (first entry)
XX
DE   Amino acid sequence of a murine PCTG4 protein.
XX
KW   Human; PCTG4 region; X chromosome; q13 region; polymorphism;
KW   mental retardation; autism; depression; bipolar affective disorder;
KW   hypothyroidism; OPA gene; neuropsychiatric disorder.
XX
OS   Mus sp.
XX
PN   WO9955915-A2.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09365.
XX
```

```
PR   29-APR-1998;   98US-0083465.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA   (IOWA ) UNIV IOWA RES FOUND.
XX
PI   Philibert RA,  Ginns EI;
XX
DR   WPI; 2000-126357/11.
XX
PT   Identification of polymorphisms in the PCTG4 region of Xq13 for
PT   diagnosing mental retardation or autism -
XX
PS   Example 7; Page 81-84; 100pp; English.
XX
CC   The present sequence represents a murine PCTG4 protein. Polymorphisms
CC   in the human PCTG4 region of chromosome Xq13 are associated with
CC   mental retardation, autism, depression, bipolar affective disorder or
CC   hypothyroidism. One 12 bp insertion polymorphism occurs within the
CC   coding region of the human OPA gene, and introduces a 4 amino acid
CC   insertion in a putative OPA domain. This domain has been shown to be
CC   involved in tissue specific expression. Another polymorphism consists
CC   of a pentanucleotide repeat approximately 7 kb upstream of the 12 bp
CC   polymorphism. Another polymorphisms consists of a dinucleotide repeat
CC   approximately 4.5 kb downstream of the 12 bp polymorphism. The
CC   specification describes a method for screening for polymorphisms in a
CC   PCTG4 nucleic acid sequence obtained from a subject. The PCTG4 related
CC   sequences within the q13 region of the X chromosome have polymorphisms
CC   associated with neuropsychiatric disorders.  The methods can be used to
CC   screen for the presence of a heritably linked form of mental retardation,
CC   autism, depression, bipolar affective disorder or hypothyroidism.
XX
SQ   Sequence   2074 AA;


  Query Match             34.0%;  Score 76.5;  DB 21;  Length 2074;
  Best Local Similarity   40.0%;  Pred. No. 1.2;
  Matches   16;  Conservative   12;  Mismatches    7;  Indels    5;  Gaps    1;

Qy       1 PEDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           ||  |:: :: ||: :||:::|| | ||:     :|:|||
Db    1938 peqqqqqqqqqqqqqqqqqqqqqqqqqq-----qqqqqq 1972


RESULT  10
Y22191
ID   Y22191 standard; Protein; 910 AA.
XX
AC   Y22191;
XX
DT   10-SEP-1999  (first entry)
XX
DE   Mouse brain CNG-1 protein sequence.
XX
KW   BCNG; brain cyclic nucleotide gated ion channel; epilepsy; hyperalgesia;
KW   Alzheimer's Disease; Parkinson's Disease; long QT syndrome; dyslexia;
KW   sick sinus syndrome; age-related memory loss; cystic fibrosis;
KW   sudden death syndrome; pacemaker rhythm dysfunction; sensory disorder;
KW   auditory disorder; respiratory disorder; attention deficit disorder;
KW   learning disability; drug addiction; therapy; mBCNG-1.
XX
OS   Mus sp.
XX
PN   WO9932615-A1.
XX
PD   01-JUL-1999.
XX
PF   23-DEC-1998;   98WO-US27630.
XX
PR   28-MAY-1998;   98US-0086436.
PR   23-DEC-1997;   97US-0997685.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Bartsch D,  Grant S,  Kandel ER,  Santoro B,  Siegelbaum S;
PI   Tibbs G;
XX
DR   WPI; 1999-418922/35.
DR   N-PSDB; X84442.
XX
PT   An isolated nucleic acid encoding a brain or heart cyclic
PT   nucleotide-gated ion channel
XX
PS   Claim 16; Page 185-188; 213pp; English.
XX
CC   This sequence is the brain cyclic nucleotide-gated ion channel
CC   (BCNG) of the invention, designated mBCNG-1. BCNG and BCNG-related
CC   proteins are useful in screening for compounds that modulate, interact or
CC   affect expression. Compounds, e.g. antagonists and agonists, identified
CC   in the methods are useful for modulating BCNG or BCNG-related protein
CC   activity. Modulation is increased or decreased ion permissivity or ion
CC   flow rate. Modulators of BCNG can be used to treat a neurological, renal,
CC   pulmonary, hepatic or cardiovascular condition. Such conditions include
CC   epilepsy, Alzheimer's Disease, Parkinson's Disease, long QT syndrome,
CC   sick sinus syndrome, age-related memory loss, cystic fibrosis, sudden
CC   death syndrome or pacemaker rhythm dysfunction. BCNG or BCNG-related
CC   protein can also be used to treat sensory disorders, e.g. blindness, loss
CC   of vision, loss of smell, numbness and lack of ability to taste. Also
CC   treatable are auditory disorders, respiratory disorders due to defects in
CC   central nervous system areas that control respiration or defects in the
CC   lungs, dyslexia, attention deficit disorder or learning disabilities,
CC   drug addiction and regulation of cell secretions. The proteins are useful
CC   targets for screening for drugs that are effective in the control of pain
CC   and hyperalgesia.
XX
SQ   Sequence   910 AA;


  Query Match             33.3%;  Score 75;  DB 20;  Length 910;
  Best Local Similarity   37.8%;  Pred. No. 0.77;
  Matches   14;  Conservative   13;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       4 PQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
           || : :: | : :||:::|| | ||:  :: :|:|||
Db     727 pqsqvqqtqtqtqqqqqqqqqqqqqqqqqqqqqqq 763


RESULT  11
W95073
ID   W95073 standard; Protein; 86 AA.
XX
AC   W95073;
XX
DT   20-MAY-1999  (first entry)
XX
DE   GST-HD fusion protein GST-HD51DELP.
XX
KW   Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
KW   polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;
KW   Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
KW   systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
KW   bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "this residue is connected to a GST protein
FT                           which is not indicated in the sequence"
XX
PN   WO9906838-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-EP04810.
XX
```

PR 01-AUG-1997; 97EP-0113320.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
XX
DR WPI; 1999-153955/13.
XX
PT Detecting amyloid-like fibrils or protein aggregates insoluble in
PT detergent or urea - from their retention on a filter, used for
PT diagnosis, particularly of diseases associated with polyglutamine
PT expansion
XX
PS Disclosure; Fig 8; 56pp; English.
XX
CC The invention relates to the detection of amyloid-like fibrils or protein
CC aggregates, insoluble in detergents or urea. The method comprises: (a)
CC applying material suspected of containing protein aggregates to a filter;
CC and (b) detecting retention of protein aggregates on the filter. This
CC method also helps to identify inhibitors of protein aggregates formation.
CC The method is particularly used to detect protein aggregates that are
CC indicative of disease, for assessing onset or progression of the
CC diseases. The inhibitors identified are potential therapeutic agents for
CC treating the diseases. Other applications include detection of inclusion
CC bodies in bacteria and to study kinetics of aggregate formation. Diseases
CC associated with polyglutamine expansion are particularly diagnosed, e.g.
CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
CC scrapie. The protein aggregates can now be detected simply, routinely and
CC rapidly, without requiring sophisticated equipment. The method can be
CC made quantitative, by analysing a series of dilutions, and can be
CC automated to allow many samples to be analysed on the same filter.
CC Sequences W95072-75 represent GST-HD fusion proteins.
XX
SQ Sequence 86 AA;

Query Match 32.4%; Score 73; DB 20; Length 86;
Best Local Similarity 35.9%; Pred. No. 0.14;
Matches 14; Conservative 15; Mismatches 10; Indels 0; Gaps 0;

Qy 2 EDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
: |:: :: ||: :||:::|| | ||: :: :|:|||
Db 25 qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq 63

RESULT 12
W95078
ID W95078 standard; Protein; 86 AA.
XX
AC W95078;
XX
DT 20-MAY-1999 (first entry)
XX
DE GST-HD fusion protein GST-HD51DELP.
XX
KW Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;
KW protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;
KW Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;
KW dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;
KW GST-HD; HD.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "this residue is connected to a GST protein
FT which is not indicated in the sequence"
XX
PN WO9906545-A2.
XX

PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-EP04811.
XX
PR 01-AUG-1997; 97EP-0113306.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
XX
DR WPI; 1999-153775/13.
XX
PT Composition containing fusion protein that includes amyloidogenic
PT peptide - able to self-assemble into fibrils or aggregates, used to
PT detect and monitor neuronal diseases, and also to screen for
PT therapeutic inhibitors
XX
PS Disclosure; Fig 8; 62pp; English.
XX
CC The invention relates to a composition comprising a fusion protein of (i)
CC (poly)peptide that increases solubility and/or prevents aggregation of
CC fusion protein, and (ii) amyloidogenic (poly)peptide that can self-
CC assemble into amyloid-like fibrils or protein aggregates. Host cells
CC transformed with a vector containing the nucleic acid encoding the fusion
CC protein are used for the recombinant expression of the fusion protein.
CC The composition is used to detect onset and progression of diseases
CC associated with fibrils/protein aggregates. It is potentially useful for
CC treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-
CC repeat expansion conditions such as Huntington's disease (HD), spinal and
CC bulbar muscular atrophy, dentatorubral pallidoluysian atrophy,
CC spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on
CC release of the amyloidogenic polypeptide from fusion protein have a
CC precise starting time for aggregate formation, allowing kinetic
CC measurements, and use of an enzyme for cleavage allows testing under
CC physiological conditions. Sequences W95077-80 represent GST-HD fusion
CC proteins.
XX
SQ Sequence 86 AA;

Query Match 32.4%; Score 73; DB 20; Length 86;
Best Local Similarity 35.9%; Pred. No. 0.14;
Matches 14; Conservative 15; Mismatches 10; Indels 0; Gaps 0;

Qy 2 EDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
: |:: :: ||: :||:::|| | ||: :: :|:|||
Db 25 qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq 63

RESULT 13
W95075
ID W95075 standard; Protein; 94 AA.
XX
AC W95075;
XX
DT 20-MAY-1999 (first entry)
XX
DE GST-HD fusion protein GST-HD51DELPBio.
XX
KW Amyloid-like fibril; protein aggregate; inhibitor; inclusion body;
KW polyglutamine expansion; Huntington's disease; Alzheimer's disease; HD;
KW Parkinson's disease; spinal; bulbar muscular atrophy; type II diabetes;
KW systemic amyloidosis; spinocerebellar ataxia; kuru; familial insomnia;
KW bovine spongiform encephalopathy; kuru; scrapie; GST-HD; fusion protein.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "this residue is connected to a GST protein
FT which is not indicated in the sequence"
XX

PN WO9906838-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-EP04810.
XX
PR 01-AUG-1997; 97EP-0113320.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
XX
DR WPI; 1999-153955/13.
XX
PT Detecting amyloid-like fibrils or protein aggregates insoluble in
PT detergent or urea - from their retention on a filter, used for
PT diagnosis, particularly of diseases associated with polyglutamine
PT expansion
XX
PS Disclosure; Fig 8; 56pp; English.
XX
CC The invention relates to the detection of amyloid-like fibrils or protein
CC aggregates, insoluble in detergents or urea. The method comprises: (a)
CC applying material suspected of containing protein aggregates to a filter;
CC and (b) detecting retention of protein aggregates on the filter. This
CC method also helps to identify inhibitors of protein aggregates formation.
CC The method is particularly used to detect protein aggregates that are
CC indicative of disease, for assessing onset or progression of the
CC diseases. The inhibitors identified are potential therapeutic agents for
CC treating the diseases. Other applications include detection of inclusion
CC bodies in bacteria and to study kinetics of aggregate formation. Diseases
CC associated with polyglutamine expansion are particularly diagnosed, e.g.
CC Huntington's, Alzheimer's or Parkinson's diseases; spinal and bulbar
CC muscular atrophy; spinocerebellar ataxia; systemic amyloidosis; type II
CC diabetes; bovine spongiform encephalopathy; kuru; familial insomnia;
CC scrapie. The protein aggregates can now be detected simply, routinely and
CC rapidly, without requiring sophisticated equipment. The method can be
CC made quantitative, by analysing a series of dilutions, and can be
CC automated to allow many samples to be analysed on the same filter.
CC Sequences W95072-75 represent GST-HD fusion proteins.
XX
SQ Sequence 94 AA;

Query Match 32.4%; Score 73; DB 20; Length 94;
Best Local Similarity 35.9%; Pred. No. 0.15;
Matches 14; Conservative 15; Mismatches 10; Indels 0; Gaps 0;

Qy 2 EDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
: |:: :: ||: :||:::|| | ||: :: :|:|||
Db 25 qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq 63

RESULT 14
W95080
ID W95080 standard; Protein; 94 AA.
XX
AC W95080;
XX
DT 20-MAY-1999 (first entry)
XX
DE GST-HD fusion protein GST-HD51DELPBio.
XX
KW Fusion protein; amyloidogenic polypeptide; amyloid-like fibril; scrapie;
KW protein aggregate; Alzheimer's disease; CAG-repeat expansion; spinal;
KW Huntington's disease; bulbar muscular atrophy; spinocerebellar ataxia;
KW dentatorubral pallidoluysian atrophy; Creutzfeld-Jakob disease; enzyme;
KW GST-HD; HD.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1
FT /note= "this residue is connected to a GST protein
FT which is not indicated in the sequence"
XX
PN WO9906545-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-EP04811.
XX
PR 01-AUG-1997; 97EP-0113306.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Bates G, Lehrach H, Scherzinger E, Wanker E;
XX
DR WPI; 1999-153775/13.
XX
PT Composition containing fusion protein that includes amyloidogenic
PT peptide - able to self-assemble into fibrils or aggregates, used to
PT detect and monitor neuronal diseases, and also to screen for
PT therapeutic inhibitors
XX
PS Disclosure; Fig 8; 62pp; English.
XX
CC The invention relates to a composition comprising a fusion protein of (i)
CC (poly)peptide that increases solubility and/or prevents aggregation of
CC fusion protein, and (ii) amyloidogenic (poly)peptide that can self-
CC assemble into amyloid-like fibrils or protein aggregates. Host cells
CC transformed with a vector containing the nucleic acid encoding the fusion
CC protein are used for the recombinant expression of the fusion protein.
CC The composition is used to detect onset and progression of diseases
CC associated with fibrils/protein aggregates. It is potentially useful for
CC treatment of such diseases (e.g. Alzheimer's disease, scrapie or CAG-
CC repeat expansion conditions such as Huntington's disease (HD), spinal and
CC bulbar muscular atrophy, dentatorubral pallidoluysian atrophy,
CC spinocerebellar ataxia, Creutzfeld-Jakob disease). Assay methods based on
CC release of the amyloidogenic polypeptide from fusion protein have a
CC precise starting time for aggregate formation, allowing kinetic
CC measurements, and use of an enzyme for cleavage allows testing under
CC physiological conditions. Sequences W95077-80 represent GST-HD fusion
CC proteins.
XX
SQ Sequence 94 AA;

Query Match 32.4%; Score 73; DB 20; Length 94;
Best Local Similarity 35.9%; Pred. No. 0.15;
Matches 14; Conservative 15; Mismatches 10; Indels 0; Gaps 0;

Qy 2 EDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
: |:: :: ||: :||:::|| | ||: :: :|:|||
Db 25 qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq 63

RESULT 15
Y58633
ID Y58633 standard; Protein; 1299 AA.
XX
AC Y58633;
XX
DT 11-APR-2000 (first entry)
XX
DE Protein regulating gene expression PRGE-26.
XX
KW Protein regulating gene expression; PRGE-26; human;
KW cell proliferation; antiproliferative; inflammation;
KW antiinflammatory; therapy; diagnosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 85

```
FT                   /note= "O-phosphorylated"
FT   Modified-site   139
FT                   /note= "O-phosphorylated"
FT   Modified-site   160
FT                   /note= "O-phosphorylated"
FT   Modified-site   199
FT                   /note= "O-phosphorylated"
FT   Modified-site   225
FT                   /note= "O-phosphorylated"
FT   Modified-site   277
FT                   /note= "O-phosphorylated"
FT   Modified-site   284
FT                   /note= "O-phosphorylated"
FT   Modified-site   335
FT                   /note= "O-phosphorylated"
FT   Modified-site   340
FT                   /note= "O-phosphorylated"
FT   Modified-site   382
FT                   /note= "O-phosphorylated"
FT   Modified-site   392
FT                   /note= "O-phosphorylated"
FT   Modified-site   412
FT                   /note= "O-phosphorylated"
FT   Modified-site   424
FT                   /note= "O-phosphorylated"
FT   Modified-site   480
FT                   /note= "O-phosphorylated"
FT   Modified-site   505
FT                   /note= "O-phosphorylated"
FT   Modified-site   684
FT                   /note= "O-phosphorylated"
FT   Modified-site   690
FT                   /note= "O-phosphorylated"
FT   Modified-site   702
FT                   /note= "O-phosphorylated"
FT   Modified-site   708
FT                   /note= "O-phosphorylated"
FT   Modified-site   945
FT                   /note= "O-phosphorylated"
FT   Modified-site   1001
FT                   /note= "O-phosphorylated"
FT   Modified-site   1067
FT                   /note= "O-phosphorylated"
FT   Modified-site   1166
FT                   /note= "O-phosphorylated"
FT   Modified-site   1253
FT                   /note= "O-phosphorylated"
FT   Modified-site   1295
FT                   /note= "O-phosphorylated"
FT   Modified-site   105
FT                   /note= "O-phosphorylated"
FT   Modified-site   122
FT                   /note= "O-phosphorylated"
FT   Modified-site   225
FT                   /note= "O-phosphorylated"
FT   Modified-site   308
FT                   /note= "O-phosphorylated"
FT   Modified-site   351
FT                   /note= "O-phosphorylated"
FT   Modified-site   359
FT                   /note= "O-phosphorylated"
FT   Modified-site   377
FT                   /note= "O-phosphorylated"
FT   Modified-site   405
FT                   /note= "O-phosphorylated"
FT   Modified-site   461
FT                   /note= "O-phosphorylated"
FT   Modified-site   665
FT                   /note= "O-phosphorylated"
FT   Modified-site   998
FT                   /note= "O-phosphorylated"
FT   Modified-site   1092
FT                   /note= "O-phosphorylated"
FT   Modified-site   1108
FT                   /note= "O-phosphorylated"
FT   Modified-site   1260
FT                   /note= "O-phosphorylated"
FT   Modified-site   1285
FT                   /note= "O-phosphorylated"
FT   Modified-site   43
FT                   /note= "O-phosphorylated"
FT   Modified-site   136
FT                   /note= "O-phosphorylated"
FT   Modified-site   268
FT                   /note= "O-phosphorylated"
FT   Modified-site   168
FT                   /note= "O-phosphorylated"
FT   Modified-site   91
FT                   /note= "N-glycosylated"
FT   Modified-site   934
FT                   /note= "N-glycosylated"
FT   Modified-site   991
FT                   /note= "N-glycosylated"
FT   Modified-site   1031
FT                   /note= "N-glycosylated"
FT   Modified-site   1090
FT                   /note= "N-glycosylated"
FT   Modified-site   1098
FT                   /note= "N-glycosylated"
FT   Modified-site   1235
FT                   /note= "N-glycosylated"
FT   Modified-site   1246
FT                   /note= "N-glycosylated"
FT   Modified-site   1282
FT                   /note= "N-glycosylated"
XX
PN   WO9964596-A2.
XX
PD   16-DEC-1999.
XX
PF   11-JUN-1999;   99WO-US13281.
XX
PR   12-JUN-1998;   98US-0089029.
PR   29-JUL-1998;   98US-0094575.
PR   14-OCT-1998;   98US-0104624.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Lal P,  Yue H,  Tang YT,  Hillman JL,  Bandman O,  Corley NC;
PI   Guegler KJ,  Gorgone GA,  Baughn MR,  Patterson C,  Lu DAM;
XX
DR   WPI; 2000-116543/10.
DR   N-PSDB; Z57864.
XX
PT   New human polypeptides that regulate gene expression, for treatment,
PT   prevention and diagnosis of, e.g. cancer -
XX
PS   Claim 1; Page 110-113; 150pp; English.
XX
CC   The present sequence is that of new human protein regulating gene
CC   expression PRGE-26, deduced from Incyte clone PITUNOT01 (see Z57864)
CC   isolated from pituitary gland cDNA library.  PRGE-26 is expressed in
CC   reproductive, nervous and gastrointestinal tissues associated with
CC   cell proliferative and inflammation diseases, disorders or conditions.
CC   It is characterised as a glutamine-rich protein.  The invention
CC   provides PRGE polypeptides (see Y58608-38) and polynucleotides (see
CC   Z57839-69), expression vectors, host cells, antibodies, agonists and
CC   antagonists.  It also provides methods for diagnosing, treating or
CC   preventing disorders associated with expression of PRGE.
XX
SQ   Sequence   1299 AA;


  Query Match             32.2%;  Score 72.5;  DB 21;  Length 1299;
  Best Local Similarity   41.0%;  Pred. No. 2;
  Matches   16;  Conservative   11;  Mismatches   11;  Indels    1;  Gaps    1;
```

```
Qy      2 EDPQRRYEECQQECRQQEERQQPQCQQRCLKRFEQEQQQ 40
          |: :|: :| : : ||:|  :| | ||  |:| :|:|||
Db    878 eeeerkrkelevq-rqkelmrqrqqqqealrrlqqqqqq 915

Search completed: March  1, 2001, 15:47:23
Job time: 248 sec
```